Statistically Accurate Sensor Networking

C. M. Okino
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109

M. G. Corr
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025-3493

Abstract—**In this paper, we propose an alternate approach to ad-hoc networking called** *Best Effort multi-Hop Geographical Routing* **(BEHGR). BEHGR does not fit under the current classifications of on-demand or table-based approaches to ad-hoc routing, but instead statistically attempts to dynamically route packets to a central location in a "best effort" manner. The basis of such a protocol assumes that a sufficient measure of the performance of the network is a statistically accurate representation of the overall collected sensor data. The metrics representing the performance include the concept of** *currentness*.

I. Introduction

In recent years, considerable work has been placed toward analyzing and presenting innovative protocols for ad-hoc networking [1] [2], where the overlying assumption is a direct application of these protocols toward sensor networks. In this paper, we identify some of the specific issues unique to types of sensor networks that allow for potentially new directions in protocols of distributed sensor networking. Initially, we review key aspects of the current trends in ad-hoc networking and then leverage assumptions specific to some sensor networks. We propose a computationally relaxed approach to distributed multi-hop data collection and present some results.

A. Overview of ad-hoc networking aspects

In the so-called infrastructureless mobile network environment, there are no fixed routers, all nodes may have mobility, and connectivity is typically dynamic. The current approach is to classify Ad-hoc routing protocols either as table-driven or on-demand [2]. Table driven protocols such as DSDV, CGSR, and WRP attempt to maintain up-to-date routing information from each node to all other nodes through routing tables. In source-initiated on-demand protocols such as AODV, DSR, TORA, ABR SSR, a node will request a route discovery process, establish a route, and maintain the route to a destination until the route is no longer accessible or the route is no longer required. For both approaches, resources are allocated for location awareness (at least for the initial state) and some form of route topology mapping.

In a distributed sensor network, the criteria for maintaining a functional network may differ from these recently proposed ad-hoc routing protocols. Specifically, the act of collecting data does not necessarily require connectivity between the nodes, but rather, sufficient connectivity to at least one node used to route information toward some central location which

we shall call *home*. Moreover, the ultimate objective of the sensor module is to gather or sense a measurement. Processing and forwarding data is secondary but still needed. Last, collection of most recent information of the entire (or partial) sensor network need only be statistically accurate. Specifically, we propose a global class of protocols developed for *statistically accurate sensor networks* (SASN), or rather a network capable of statistically representing the currentness [3] of the distributed information. Thus, in a SASN, it is sufficient that only some of the collected and transmitted data from each of the nodes reach home in order to provide an accurate picture of the measured and collected data.

Geographical routing methodologies utilizing GPS have been presented for position identification [4] [5] [6] [7] [8]. Extending on this concept, [9] utilizes position information explicitly as the sensor identifier as oppose to traditional IP labeling or the wireless equivalent of layer 1 external gateway routing protocols such as BGP-4, thus eliminating the need to append position information in the data payload. Moreover, in a dynamic environment, minimal interaction of link connection status while maintaining relative position within in the network is sufficient for statistically forwarding information toward a central location, i.e. geographic awareness and the "position database" concept used for node position identification [10] is not a requirement for distributed sensor measurement networking.

In the following sections, we characterize SASN's and then propose an approach synonymous with the original Internet Protocol philosophy of "best effort" networking while utilizing the minimal knowledge of geographical positioning.

II. Statistically Accurate Sensor Networks

A statistically accurate sensor network is a network such that the database collection of information from the sensors provides a sufficiently accurate representation of the distributed nature of the sensor network. In SASN's, no routing tables are required and no route discovery procedure is explicitly executed end-to-end. Nodes either act as clients to forward packets or as server's in order to receive packets. At first, the behavior for servers and clients may seem reversed to the terminology server and client respectively, but from the perspective of establishing a connection in a point to multi-point network using RF transceivers and forwarding packets toward a central location, the terminology holds. Relative position towards a central location such as *home* determines if a packet is in a loop or is effectively being handed off to a node closer to the central location, i.e. intended unidirectional flow of sensor

information. Error checking is performed, but there are no acknowledgments and no guarantees that the server has properly received the packet forwarded from the client. One could also view this as a form of connectionless oriented location aware ad-hoc sensor networking [11].

We assume we have n sensors (nodes), where each node is capable of uni-directional transmission based on power level of radius d_{RF} over a unit circular area. Gupta and Kumar [12] have shown that given that each node covers RF circular transmission area $\pi d_{RF}^2 = \frac{\log n + c(n)}{n}$, then the network approaches connectivity with probability 1. The connectivity requirement stated in [12] is in fact stronger than necessary for our network. We shall view this as a sufficient condition and leave the precision of connectivity for future work.

A. Performance Metric

The traditional approach of throughput and delay remain significant in a statistically accurate sensor network. We consider *throughput* as the total number of packets that arrived up to time t divided by the total number of packets generated up to time t. We can then consider the throughput of a given node within the network or the overall throughput of the network. The network we consider may potentially have high loss in the sense of dropped packets due to queue overflow or dropped packets due to bit errors in transmission over a noisy channel.

Another metric in addition to throughput is providing a measure of delay. We attempt to relax the condition of delay to allow for a metric of being sufficiently current information. In general, one could view currentness as a probabilistic form of delay.

We propose a new metric of performance for ad-hoc sensor networks called currentness. The concept of currentness was first formally proposed for estimating the speed of re-indexing web server webpage access [3]. We refine and formalize the definition for ad-hoc sensor networks.

Let $\delta(i)$ be the interval of time between generation and arrival at the final destination of information piece i.

Definition 1 ((α, β)-currency of a source in a network)
A source in a network is said to be (α, β) current if each piece of received information i from the source arrives at the destination within some interval of time $\delta(i) \leq \beta$ with probability α.

Similarly, we can present the concept of an entire network being (α, β) current with respect to a specific destination.

Definition 2 ((α, β)-currency of a network) A network is said to be (α, β) current if each network generated piece of received information i arrives within some interval of time $\delta(i) \leq \beta$ with probability α.

Note that this concept allows for a relaxed approach to receiving up-to-date information while maintaining a level of performance.

B. Best Effort multi-Hop Geographical Routing

We consider a multi-hop distributed sensor network where each module is a node that is randomly placed a number of geographical units apart where each geographical unit could be considered as the maximum RF reception distance. Each sensor or module is identical in nature and will adapt within the network based on the geographical position relative to some predefined home base.

Each sensor will oscillate between amount of time in client mode and amount of time in server mode. Moreover, each sensor will dynamically adjust time allocated as a client (client mode period) versus time allocated as a server (server mode period), where a client's main purpose is to forward data to the follow on node (in this case also called the server) and the server's main purpose is to receive data. In addition, each sensor allocates a fixed amount of time locally collecting sensor data.

The likelihood of a sensor being in client mode as oppose to server mode is directly dependent on its relative geographical position to home. Specifically, a sensor located near home will adapt toward statistically allocating more time as a server, while a sensor located a number of RF hops away from home will adapt toward statistically allocating more time as a client. The statistical allocation of allotted time will result in variation in the likelihood of a sensor in client mode synchronizing in time with a sensor in server mode. The relationship between duration of time as a client increasing with respect to the number of hops from a central location holds for cases of aggregating data among the hops.

The algorithm for allocation of time in server versus client mode is shown in Figure 1 where $\hat{C}(i,k)$ and $\hat{S}(i,k)$ is the amount of time allotted during oscillation round i for the kth tier module as a client and server respectively. Prior to entering the oscillation period (client and server mode), a sensor waits in client mode for the home location to broadcast and forward the current location of the home node in order to provide a
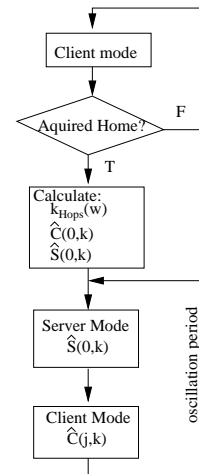

Fig. 1. Algorithm for server/client time BEHGR protocol time allocation

direction to route. The process of broadcasting the coordinates of home are repeated in a radial fashion relative to the current location of home.

Let W be the set of all nodes in the network. Define the home sensor unit as a base station located at Cartesian coordinates, $(\hat{x}, \hat{y})$. For the module w, we have coordinates (x_w, y_w), and so the radial distance of the wth sensor from home is

$$d_w = \sqrt{(x_w - \hat{x})^2 + (y_w - \hat{y})^2} \ .$$

Let

$$k_{hops}(w) = \frac{d_w}{d_{RF}} \ ,$$

be the multi-hop parameter of w, where d_{RF} is the maximum reception RF distance. For the actual implementation of an algorithm, since the true distance is dependent on bit error rates, we select a distance with low probability of a bit error, approximately 10^{-6}. Thus, $k_{hops}(w)$ provides a measure of the approximate number of hops required by sensor w to reach home assuming a chain of $k_{hops}(w)$ sensors separated at a distance of d_{RF}. Clearly, the actual number of hops will be dependent on the topology of the distributed sensor network which is is not known in an ad-hoc or dynamically changing network. We define the tier value k for module w as

$$k = \lfloor k_{hops} \rfloor \ .$$

Let the set of modules belonging to the kth tier be

$$\hat{N}(k) = \{w : \lfloor k_{hops}(w) \rfloor = k \ \forall \ w \in W\} \ ,$$

where $\lfloor x \rfloor$ is the floor function of a real number x.

Let $\hat{C}(i, k)$ and $\hat{S}(i, k)$ be the Client and Server state time duration of the kth tier modules during oscillation round i, where $i = 0$ is the initial state time duration.

For Server mode, the duration of time remaining in Server mode is fixed such that

$$\hat{S}(i, k) = H_{max} - k \quad \forall i \geq 0 \ ,$$

where H_{max} is the maximum allowed number of hops from the outermost module in the network to Home.

The Client state time duration varies in an adaptive manner, conditioned on the attempted lock to a Server w' where the tier value of Server w' is k'. We write the initial state of the kth tier modules as

$$\hat{C}(0, k) = H_{max} + k \ .$$

For all other oscillation rounds, we condition the client's following oscillation round duration based on locking to a server in the proper direction. Specifically, if Client w locks with server w', but does not receive a valid tier value, $k < k'$, then

$$\hat{C}(i, k) = \max\{H_{min}, \hat{C}(i - 1, k) - T_{step}\} \ ,$$

else if Client w locks with Server w' and receives a valid tier value, $k \geq k'$, then

$$\hat{C}(i, k) = \hat{C}(i - 1, k) \ ,$$

else if Client w does not lock at all, (k' unknown), then

$$\hat{C}(i, k) = \min\{\hat{C}(i - 1, k) + T_{step}, \gamma H_{max}\} \ ,$$

where H_{min} is the minimum number of slots allowed as a client, T_{step} is the potential variation in number of time slots for each oscillation round, and γ is a positive real number greater that 1.

As stated earlier, BEHGR does not provide and end-to-end route discovery process. However, BEHGR does implement a route discover process on a per link basis, comprising of the initial calculation of hop count and client duration update described earlier in the paper.

III. ANALYSIS

In this section we obtain some bounds on some performance metrics and plot some preliminary performance using a real test-bed of sensors.

The following lemma provides a sense of the worst case delay due to hop synchronization over multiple hops with no competing nodes at each hop and no physical link errors.

Lemma 3 (Hop Delay) For an error free environment where each hop contains a single module, the worst case number of time slots to propagate a piece of information from a module w at the kth tier is

$$T(w) = (k + 1)H_{max} + \frac{2 + k - k^2}{2}$$

time slots.

As depicted in Figure 2, utilizing the BEHGR protocol, the worst case number of time slots required to forward a packet home due to synchronization delay but with zero interference and zero queuing delay is critical to the selection of the maximum number of hops H_{max}.

Consider a $k \times k$ grid as depicted in Figure 3, where each sensor node is located on a grid such that the nearest neighbor
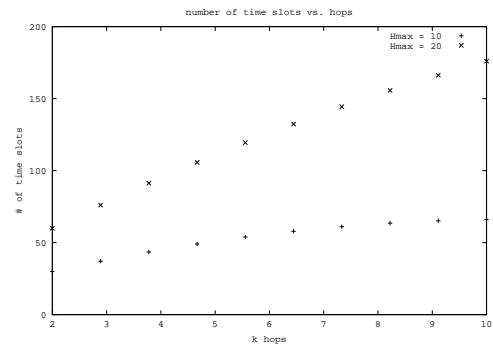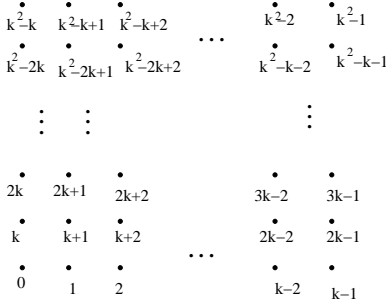

Fig. 2. This depicts the worst case optimal number of time slots required to route a packet from the kth hop with no competing nodes using BEHGR and a maximum number of hops $H_{max} = 10, 20$.

Fig. 3. $k \times k$ grid with spacing d_{RF}

is either on the horizontal or vertical axis at a distance d_{RF} such that only these nearest neighbors are within RF range of the node of interest. Suppose that Home is located in the lower left corner. Then each of the other $k^2 - 1$ nodes are expected to route information via neighboring nodes back to Home.

Theorem 4 (Currentness for $k \times k$ grid) For a $k \times k$ grid, where packets are generated at most one per oscillation round, the probability that the number of time slots is less than or equal to some value β is

$$Pr\{T_{k\times k} \le \beta\} \le 1 - \frac{1 - \frac{1}{h_{max}}}{k^2 - 1} \sum_{i=2}^{k} i(1 - \frac{1}{H_{max}})^{\frac{\beta}{i}}$$
$$- \frac{1 - \frac{1}{H_{max}}}{k^2 - 1} \sum_{i=1}^{k-1} i(1 - \frac{1}{H_{max}})^{\frac{\beta}{2k-i}} .$$

Proof of Theorem 4: Let q be the probability of a packet successfully being received in a time slot. Then the probability of a packet successfully traversing a hop in time slot i is $Pr\{T_h = i\} = (1 - q)^i q$.

Let $T_{Q,w}$ be the amount of time a packet spends in the node w queue. Let $\hat{T}_{Q,k}$ be the minimum amount of time a packet spends in the kth tier queue. Thus, we have $\hat{T}_{Q,k} \le T_{Q,w}$.

Let T_d be the amount of time required to traverse d hops including queuing delay. Then, we have

$$Pr\{T_d \le \beta\} = Pr\{d(T_h + \hat{T}_{Q,k}) \le \beta\}$$
$$= Pr\{T_h \le \frac{\beta - d\hat{T}_{Q,k}}{d}\}$$
$$= \sum_{i=0}^{\frac{\beta - d\hat{T}_{Q,k}}{d}} (1 - q)^i q$$
$$= q \frac{1 - (1 - q)^{\frac{\beta - d\hat{T}_{Q,k}}{d} + 1}}{1 - (1 - q)}$$
$$= 1 - (1 - q)^{\frac{\beta - d\hat{T}_{Q,k}}{d} + 1}$$
$$\le 1 - (1 - q)^{\frac{\beta}{d} + 1} .$$

Let $\hat{p}_d = 1 - (1 - q)^{\frac{\beta}{d} + 1}$. Assume that Home is located in a corner. Drawing diagonal lines through nodes such that these lines are perpendicular to the diagonal line passing through

Home designates the depth of each set of nodes, where the set of nodes contained along the diagonal line closest to the Home is of depth 1 and so on until the final node on the opposite corner containing depth $2k - 2$. Thus, for a $k \times k$ grid with Home located at a corner, we have

$$Pr\{T_{k\times k} \le \beta\} \le \frac{1}{k^2 - 1} \left\{ \sum_{i=2}^{k} i\hat{p}_{i-1} + \sum_{i=1}^{k-1} i\hat{p}_{2k-i} \right\}$$
$$= \frac{1}{k^2 - 1} \left\{ \frac{2k^2 - 2}{2} - \sum_{i=2}^{k} i(1 - q)^{\frac{\beta}{i} + 1} \right\}$$
$$- \frac{1}{k^2 - 1} \sum_{i=1}^{k-1} i(1 - q)^{\frac{\beta}{2k-i} + 1}$$
$$\le \frac{\left\{ \frac{2k^2 - 2}{2} - (1 - p) \sum_{i=2}^{k} i(1 - p)^{\frac{\beta}{i}} \right\}}{k^2 - 1}$$
$$- \frac{1 - p}{k^2 - 1} \sum_{i=1}^{k-1} i(1 - p)^{\frac{\beta}{2k-i}} .$$

If $\hat{S} = 0$, we are in client mode all the time, and so $\hat{C} = H_{max} + k$. Thus, at the least, $\hat{C} \ge H_{max}$. If packets are generated at most one packet per oscillation round, then we have $p \le \frac{1}{H_{max}}$, and so we have

$$Pr\{T_{k\times k} \le \beta\} \le 1 - \frac{1 - \frac{1}{H_{max}}}{k^2 - 1} \sum_{i=2}^{k} i(1 - \frac{1}{H_{max}})^{\frac{\beta}{i}}$$
$$- \frac{1 - \frac{1}{H_{max}}}{k^2 - 1} \sum_{i=1}^{k-1} i(1 - \frac{1}{H_{max}})^{\frac{\beta}{2k-i}} ,$$

and we are done. $\square$
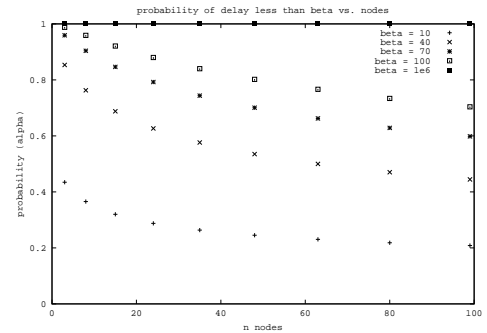
Fig. 4. Upper bound currentness profile α versus number of nodes in a $k \times k$ grid for various values of β

In Figure 4, we have the probability α of a packet arriving from a node in the network within β time slots for various number of nodes in the grid where $H_{max} = 10$. Note that Theorem 4 does not consider the queuing delay and so for large β, such as shown for $\beta = 10^6$, the bound is weak.

A. Preliminary real test-bed results

The real test-bed consists of a number of sensor modules distributed in a $k \times k$ grid arrangement as depicted in Fig-

ure 3 where each sensor node is approximately spaced d_{RF} apart. For most test cases, we assume the Home unit is in the lower left corner, and so we examine the performance of the $n = k^2 - 1$ nodes routing information toward the Home node. Each sensor module contains an 8051 microcontroller, a Motorola GPS unit, a Frequency Hop transceiver, and a Dallas semiconductor iButton temperature sensor as shown in Figure 5.
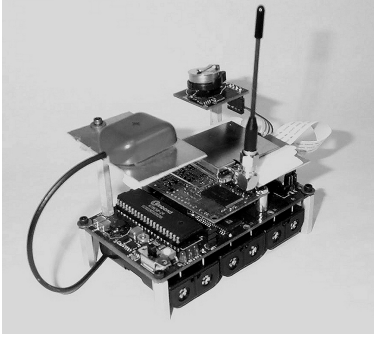

Fig. 5. Sensor module with microcontroller, GPS, RF transceiver and temperature sensor.

Tests were also performed on the relationship of currentness with respect to the positioning of Home within the grid. Specifically, we positioned Home on the corner, on an edge and in the center (odd sized grids). Preliminary test indicated that the α varied on the order of 0.1 for a $\beta \approx 10^6$. As the grid
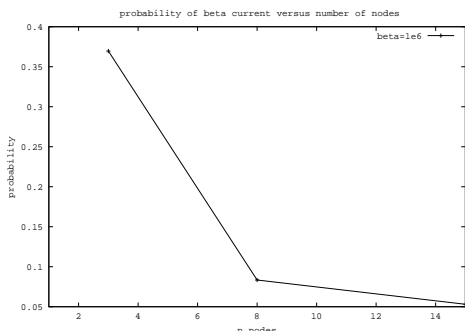

Fig. 6. Currentness versus number of nodes

size increased, overall throughput and currentness decreased rapidly. We conjecture queuing loss as the primary cause. Specifically, we believe dropped packets are the main source of error.

Throughput results indicated minimal dependency on positioning of Home. Specifically, preliminary real test-bed results have a variation on the order of 0.05%. There appears to be an inherent fairness in the algorithm in terms of allowing for balanced metrics although more tests are required to validate this claim.

In Figures 7 and 8, the percentage of packets generated with respect to the queue size has a considerable effect on the overall currentness and throughput of the network. Local data was
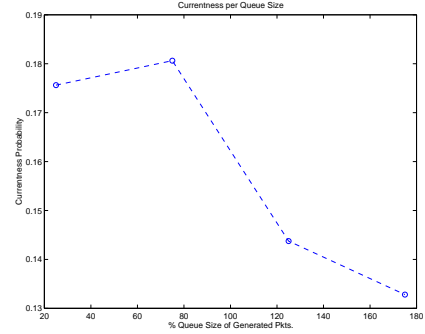

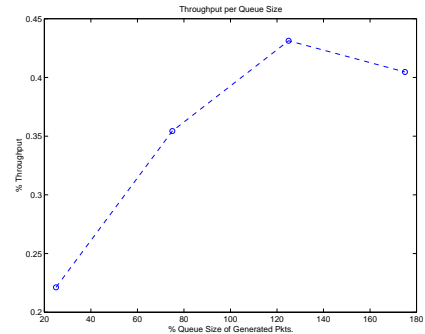Fig. 7. Currentness versus percentage of packets relative to queue size

Fig. 8. Throughput versus percentage of packets generated relative to queue size

generated at half the rate of delivery. Thus, loss due to queuing overflow was evident.

IV. DISCUSSION AND FUTURE WORK

The BEHGR protocol proposed in this paper presents an alternate approach to routing and forwarding data with minimal computation and network topology knowledge. We proposed and measured the performance of a sensor network utilizing a statistically sufficient statistic such as currentness.

The concept for developing statistically accurate sensor networking protocols represents a new set of ad hoc wireless networks worthy of investigation. The variation on the H_{max} maximum number of hops provides incite into the synchronization aspect. In some sense, the algorithm is providing a TDMA allocation of slots based on the maximum number of hops.

Queuing, fairness, and scalability in terms of hop statistics needs further analysis as well as the closed loop stability of synchronization.

The barrier problem associated with routing sensor information in a SASN in the opposite direction relative to a central location is alleviated but not eliminated by the assumption that each node is actively mobile. The mobility allows for the likelihood that a node will enter a position allowing for the stranded node to temporarily forward information toward home. Home is also allowed to move but requires some time for the network to adapt and settle to the change in position. In the case where Home moves, a broadcast message of Home's new location is

sufficient in terms of information transfer. In terms of computations, a new tier count calculation is required for each node.

Energy efficiency for battery powered sensor modules needs to be addressed in future work. Of particular interest is the work on extending the life of a battery powered network of nodes by sending remaining energy values along with a tier count as a means of determining if a server client relationship with a higher energy level can be obtained before forwarding a packet which is in the spirit of [13].

REFERENCES

[1] C. Perkins, *Ad-Hoc Networking*, Addison-Wesley, 2001.

[2] Elizabeth Royer and C-K Toh, "A review of current routing protocols for ad-hoc mobile wireless networks," in *IEEE Personal Communications Magazine*, april 1999.

[3] Brian Brewington, *Observation of changing information sources*, Phd thesis, Dartmouth College, June 2000.

[4] Y.-B. Ko and N.H. Vaidya, "Location-aided routing (lar) in mobile ad hoc networks," in *In ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM)*, November 1998.

[5] R. Morris and B. Karp, "Greedy perimeter state routing (gpsr)," *Draft, http://www.eecs.harvard.edu/ karp/gpsr4-98.ps*, apr 1998.

[6] J.C. Navas and T. Imielinski, "Geocast - geographic addressing and routing," in *In ACM/IEEE Int. Conf. on Mobile Computing and Networking (MobiCom'97)*, September 1997.

[7] S. Basagni, I. Chlamtac, V.R. Syrotiuk, and B.A. Woodward, "A distance routing effect algorithm for mobility (dream)," in *In ACM/IEEE Int. Conf. on Mobile Computing and Networking*, October 1998, pp. 76–84.

[8] K.N. Amouris, S. Papavassiliou, and Li Miao, "A position-based multizone routing protocol for wide area mobile ad-hoc networks," in *IEEE Vehicular Technology Conference 1999*, 1999, vol. 2, pp. 1365–1369.

[9] Michael G. Corr and C. Okino, "Networking reconfigurable smart sensors," in *In Proceedings of SPIE, Enabling Technologies for Law Enforcement and Security*, November 2000.

[10] Basagni S., I. Chlamtac, and V. R. Syrotiuk, "Geographic messaging in wireless ad hoc networks," in *Proceedings of the IEEE 49th Annual International Vehicular Technology Conference (VTC'99), Houston, TX,May 16-20, 1999*, 1999, vol. 3, pp. 1957–1961.

[11] Michael G. Corr, "Geographic based ad-hoc routing for distributed sensor networks," June 2001.

[12] Piyush Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming. Edited by W.M.McEneany, G. Yin, and Q. Zhang*, pp. 547–566. Birkhauser, Boston, 1998.

[13] Qun Li, J. Aslam, and D. Rus, "Online power-aware routing in wireless ad-hoc networks," in *Proceedings of ACM 7th Annual International Conference on Mobile Computing and Networking (MOBICOM)*, July 2001.